Journal of Computational Mathematics, Vol.9, No.4, 1991, 378–387.

POLYNOMIAL ACCELERATION METHODS FOR SOLVING SINGULAR SYSTEMS OF LINEAR EQUATIONS*1)

Cao Zhi-hao
(*Fudan University, Shanghai, China*)

Abstract

In this paper we study the polynomial acceleration methods for solving singular linear systems. We establish iterative schemes, show their convergence and find iteration error bounds.

§1. Introduction

For many practical problems, such as Neumann problems and those for elastic bodies with three surfaces and Poisson's equation on a sphere and with periodic boundary conditions, their finite difference and finite element formulations lead to singular but consistent systems of linear equations. In addition, when an eigenvalue problem is solved by a relaxation method, the solution of a singular linear system is involved[9]. However, as pointed in [1], methods for solving singular systems of linear equations have unfortunately been somewhat neglected in literature. Perhaps this is due to some of the difficulties involved in establishing criteria for convergence.

In this paper we study polynomial acceleration methods for solving singular linear systems. We establish iterative schemes, show their convergence and find iteration error bounds.

For convenience, we discuss real systems. All results obtained in this paper can be easily generalized to complex systems.

We use the following notations: E^n is an n-dimensional real vector space, $E^{n\times n}$ stands for a set of all $n\times n$ real matrices, $\mathcal{N}(A)$ and $\mathcal{R}(A)$ represent null space and column space (range of value) of matrix A, respectively, $\sigma(A)$ stands for the set of all eigenvalues of matrix A and A^T and A^+ are the transpose and the Moore-Penrose inverse of matrix A, respectively.

§2. Basic Iterative Methods

Consider a linear system

$$Ax = b, \tag{2.1}$$

where $A\in E^{n\times n}, x\in E^n$ and $b\in\mathcal{R}(A)$. We construct a (linear stationary) basic iterative method

$$x^{\nu+1} = Tx^{\nu} + g, \tag{2.2}$$

* Received October 31, 1988.

1) The Project Supported by the National Natural Science Foundation of China and by the Doctorate Conferrer Institutions Foundation of China.

where the iterative matrix $T \in E^{n\times n}, g \in E^n, x^\nu, x^{\nu+1} \in E^n$. (2.2) can be written as follows:

$$x^{\nu+1} = x^\nu - H(Ax^\nu - b), \tag{2.3}$$

where $H \in E^{n\times n}$. From (2.2) and (2.3) we have

$$T = I - HA, \quad g = Hb. \tag{2.4}$$

Let $x^*(x^0)$ denote a solution of (2.1) which is a limit of a vector sequence produced by an iterative method (not necessarily a linear stationary iterative method) with x^0 as an initial iterative vector. Then we define the set of error vectors U[8]:

$$U = \{y : y = x - x^*(x), \quad x \in U_0\}, \tag{2.5}$$

where U_0 is a set of the initial iterative vectors. When U is a subspace of E^n, we use $\|\cdot\|_U$ to denote a vector norm in U and the induced matrix norm. Then we use $R_\infty(\tilde{T})$,

$$R_\infty(\tilde{T}) = -\lim_{\nu\to\infty} \frac{1}{\nu} \ln \|\tilde{T}^{(\nu)}\|_U, \tag{2.6}$$

to denote the asymptotic rate of convergence of an iterative method for solving singular systems, where $\tilde{T}^{(\nu)}$ stands for the error transition operator of the ν-th iteration[8].

If we consider the linear stationary iterative method (2.2)–(2.4) and introduce the sub-spectral radius of the iterative matrix T:

$$\gamma(T) = \max\{|\lambda| : \lambda \in \sigma(T) \cup \{0\}\backslash\{1\}\}, \tag{2.7}$$

then we have the following result:

Theorem 2.1[8]. *The linear stationary iterative method* (2.2)–(2.4) *is convergent in* $U_0 \equiv E^n$ *if and only if*

(i) $\gamma(T) < 1$,

(ii) rank $(I - T) =$ rank $(I - T)^2$,

(iii) $\mathcal{N}(A) = \mathcal{N}(HA)$ *or, equivalently,* $\mathcal{N}(H) \cap \mathcal{R}(A) = \{0\}$.

When the linear stationary iterative method is convergent, U *(cf.* (2.5)) *must be a subspace:* $U = \mathcal{R}(HA)$, *and the asymptotic rate of convergence is*

$$R_\infty(T) = -\ln \gamma(T). \tag{2.8}$$

Note that the basic iterative method (2.2)–(2.4) can be derived from the splitting of matrix A:

$$A = H^+ - H^+T. \tag{2.9}$$

Definition 2.1. *The iterative method* (2.2) *is symmetrizable if for some nonsingular matrix* W *the matrix* $W(I - T)W^{-1}$ *is symmetric positive semidefinite (SPSD). Such a matrix* W *is called a symmetrization matrix.*

Obviously, if the iterative method (2.2) is symmetrizable, then the eigenvalues of T are real and matrix T is diagonalizable. Hence the condition (ii) of Theorem 2.1 is satisfied. Let

$$m(T) = \min\{\lambda : \lambda \in \sigma(T)\}, \quad M(T) = \max\{\lambda : \lambda \in \sigma(T)\};$$

then we have

$$M(T) \le 1. \tag{2.10}$$

We now consider the simplest acceleration iterative method. Let us construct an extrapolation method based on (2.2):

$$x^{\nu+1} = \omega(Tx^{\nu} + g) + (1-\omega)x^{\nu},$$

where ω is a real parameter. The above expression can be rewritten as follows:

$$x^{\nu+1} = [\omega T + (1-\omega)I]x^{\nu} + \omega g \equiv T_\omega x^{\nu} + \omega g. \tag{2.11}$$

Suppose (2.2) is symmetrizable and determine $\omega = \tilde{\omega}$ such that

$$\gamma(T_{\tilde{\omega}}) = \min_{\omega} \gamma(T_\omega). \tag{2.12}$$

Let

$$\begin{aligned} M_\gamma(T) &= \max\{\lambda : \ \lambda \in \sigma(T)\backslash\{1\}\}, \\ m_\gamma(T) &= \min\{\lambda : \ \lambda \in \sigma(T)\backslash\{1\}\}. \end{aligned} \tag{2.13}$$

Obviously, we have $M_\gamma(T) < 1$. It is not difficult to get the solution of problem (2.12):

$$\begin{aligned} &\tilde{\omega} = 2/[2 - M_\gamma(T) - m_\gamma(T)], \\ &\gamma(T_{\tilde{\omega}}) \equiv \tilde{\sigma} = [M_\gamma(T) - m_\gamma(T)]/[2 - M_\gamma(T) - m_\gamma(T)] < 1. \end{aligned} \tag{2.14}$$

Since $\gamma(T_{\tilde{\omega}}) < 1$, the optimal extrapolation method (2.11) based on (2.2) satisfies conditions (i) and (ii) of Theorem 2.1. Therefore, if the basic iterative method (2.2) is symmetrizable and condition (iii) of Theorem 2.1 is satisfied, then the optimal extrapolation method based on (2.2) (which is not necessarily convergent) is convergent.

§3. Polynomial Acceleration Based on the Symmetrizable Basic Iterative Method

In this section we always assume that the basic iterative method (2.2) is symmetrizable and a symmetrization matrix is denoted by W.

Obviously, the set of the linearly independent eigenvectors of T,

$$\{\varphi_j\}, \quad j = 1, 2, \cdots, n, \tag{3.1}$$

includes a basis for E^n. We assume the following normalization conditions are satisfied for the set of eigenvectors:

$$(W^T W\varphi_i, \varphi_j) = (W\varphi_i, W\varphi_j) = \delta_{ij}. \tag{3.2}$$

Without loss of generality, let $\{\varphi_i\}$, $i = 1, \cdots, m$, be the linearly independent eigenvectors associated with eigenvalue 1 (if any) of T. If condition (iii) of Theorem 2.1 is satisfied for the iterative method (2.2) (which is also assumed to be satisfied in this section), then we have

$$\mathcal{N}(A) = \mathcal{N}(HA) = \text{ span } \{\varphi_1, \cdots, \varphi_m\}, \tag{3.3}$$

and

$$E^n = \mathcal{N}(A) \overset{\perp}{\oplus} \mathcal{R}(HA), \tag{3.4}$$

where the orthogonality is in the following sense:

$$(W\varphi, W\psi) = 0, \quad \forall\varphi \in \mathcal{N}(A), \quad \forall\psi \in \mathcal{R}(HA). \tag{3.5}$$

By $\|\cdot\|_W$ we denote a vector norm in E^n defined as follows:

$$\|\varphi\|_W^2 = (W\varphi, W\varphi), \quad \forall\varphi \in E^n. \tag{3.6}$$

The induced matrix norm $\|\cdot\|_W$ is

$$\|A\|_W = \|WAW^{-1}\|_2, \quad \forall A \in E^{n\times n}. \tag{3.7}$$

The minimum solution with respect to the norm $\|\cdot\|_W$ of (2.1) is denoted by $\tilde{x}^{**}$; obviously, $\tilde{x}^{**} \in \mathcal{R}(HA)$. Let $\varepsilon^0 = x^0 - \tilde{x}^{**}$ be the error vector of the initial vector x^0. The error vector $\varepsilon^\nu = x^\nu - \tilde{x}^{**}$ of the iterative vector x^ν produced by the polynomial acceleration method based on (2.2) with respect to the sequence of polynomials $\{Q_j(\mu)\}$ can be expressed as follows[5]:

$$\varepsilon^\nu = Q_\nu(T)\varepsilon^0, \tag{3.8}$$

where $Q_\nu(\mu) \in \mathcal{P}_\nu$, and $\mathcal{P}_\nu$ stands for the set of all polynomials of degree $\le \nu$ and the value equals 1 when $\mu = 1$.

From (3.4) we know that x^0 and ε^0 have decompositions

$$x^0 = x_1^0 + x_2^0, \quad x_1^0 \in \mathcal{N}(A), \quad x_2^0 \in \mathcal{R}(HA)$$

and

$$\varepsilon^0 = \varepsilon_1^0 + \varepsilon_2^0, \quad \varepsilon_1^0 = x_1^0 \in \mathcal{N}(A), \quad \varepsilon_2^0 = x_2^0 - \tilde{x}^{**} \in \mathcal{R}(HA), \tag{3.9}$$

respectively. Hence

$$\varepsilon^\nu = Q_\nu(T)\varepsilon^0 = Q_\nu(T)\varepsilon_1^0 + Q_\nu(T)\varepsilon_2^0 = x_1^0 + Q_\nu(T)\varepsilon_2^0. \tag{3.10}$$

In view of (3.10), for the convergence of an iterative method for a singular system we should have

$$\varepsilon^\nu \to x_1^0, \quad \nu \to \infty,$$

that is

$$x^\nu \to x_1^0 + \tilde{x}^{**}, \quad \nu \to \infty. \tag{3.11}$$

Note that if we make the decomposition for any $x \in E^n$:

$$x = x_1 + x_2, \quad x_1 \in \mathcal{N}(A), \quad x_2 \in \mathcal{R}(HA),$$

then

$$x^*(x) = x_1 + \tilde{x}^{**}. \tag{3.12}$$

By taking $U_0 = E^n$, it is easy to know (cf. (2.5))

$$U = \{x_2 - \tilde{x}^{**} : x_2 \in \mathcal{R}(HA)\}. \tag{3.13}$$

We now have

$$\|Q_\nu(T)\varepsilon_2^0\|_W \le \|Q_\nu(T)\|_W \|\varepsilon_2^0\|_W = \max\{|Q_\nu(\lambda)| : \lambda \in \sigma(T)\backslash\{1\}\}\|\varepsilon_2^0\|_W. \tag{3.14}$$

In terms of (3.10)–(3.14) we can derive the Chebyshev semi-iterative method based on (2.2) for solving the singular system (2.1). Let

$$\tilde{\gamma}(Q_\nu(T)) = \max\{|Q_\nu(\mu)| : m_\gamma(T) \le \mu \le M_\gamma(T)\}. \tag{3.15}$$

Definition 3.1. *If the polynomial sequence* $\{\tilde{Q}_j(\mu)\}$ *is chosen such that*

$$\tilde{\gamma}(\tilde{Q}_\nu(T)) = \min\{\tilde{\gamma}(Q_\gamma(T)) : \ Q_\nu(\mu) \in P_\nu\}, \tag{3.16}$$

then the associated polynomial acceleration method is called Chebyshev semi-iterative method.

Since the derivation of the Chebyshev semi-iterative method for the singular systems is analogous to that for nonsingular systems, we only give the results and omit the details of the derivation.

Iterative Scheme (Chebyshev Semi-Iteration):

$$\begin{aligned} x^1 &= \tilde{\omega}(Tx^0 + g) + (1-\tilde{\omega})x^0, \\ x^{\nu+1} &= \tilde{\rho}_{\nu+1}[\tilde{\omega}(Tx^\nu + g) + (1-\tilde{\omega})x^\nu] + (1-\tilde{\rho}_{\nu+1})x^{\nu-1}, \quad \nu \ge 1, \end{aligned} \tag{3.17}$$

where $\tilde{\omega}$ is as in (2.14) and $\tilde{\rho}_\nu$ can be calculated resursively as follows:

$$\tilde{\rho}_1 = 1, \quad \tilde{\rho}_2 = (1-\tilde{\sigma}^2/2)^{-1}, \quad \tilde{\rho}_{\nu+1} = (1-\tilde{\rho}_\nu\tilde{\sigma}^2/4)^{-1}, \quad \nu \ge 2,$$

where $\tilde{\sigma}$ is as in (2.14).

In regard to the rate of the convergence of the Chebyshev semi-iterative method we have (cf. (3.10)–(3.14))

$$\|\tilde{\varepsilon}_2^\nu\|_W \equiv \|\tilde{Q}_\nu(T)\varepsilon_2^0\|_n \le [2\tilde{r}^{\nu/2}/(1+\tilde{r}^\nu)]\|\varepsilon_2^0\|_W, \tag{3.18}$$

where

$$\tilde{r} = \frac{1-\sqrt{1-\tilde{\sigma}^2}}{1+\sqrt{1-\tilde{\sigma}^2}}. \tag{3.19}$$

In terms of (3.10)–(3.14) we can also define the conjugate gradient acceleration method based on the basic iterative method (2.2) for the singular system (2.1).

Definition 3.2. If the polynomial sequence $\{\mathring{Q}_j(\mu)\}$ is chosen such that

$$\max\{|\mathring{Q}_\nu(\lambda)| : \lambda \in \sigma(T)\backslash\{1\}\} = \min\{\max\{|Q_\nu(\lambda)| : \lambda \in \sigma(T)\backslash\{1\}\} : Q_\nu \in P_\nu\}, \tag{3.20}$$

then the associated polynomial acceleration method is called conjugate gradient acceleration method.

It is easy to know that the following algorithm is also applicable to solving (2.1) when A is SPSD[6].

CG (Conjugate Gradient) **Algorithm**[5] :

$$\begin{aligned} &x^{\nu+1} = x^\nu + \alpha_\nu p^\nu, \quad p^\nu = r^\nu + \beta_\nu p^{\nu-1} (p^0 = r^0 = b - Ax^0), \\ &r^\nu = r^{\nu-1} - \alpha_{\nu-1}Ap^{\nu-1}, \quad \alpha_\nu = (p^\nu, r^\nu)/(p^\nu, Ap^\nu)[= (r^\nu, r^\nu)/(p^\nu, Ap^\nu)], \\ &\beta_\nu = -(r^\nu, Ap^{\nu-1})/(p^{\nu-1}, Ap^{\nu-1})[= \|r^\nu\|_2^2/\|r^{\nu-1}\|_2^2], \end{aligned} \tag{3.21}$$

or the equivalent three-term recurrence shceme:

$$x^{\nu+1} = \rho_{\nu+1}[\omega_{\nu+1}r^{\nu} + x^{\nu}] + (1-\rho_{\nu+1})x^{\nu-1}, \tag{3.22}$$

where

$$\omega_{\nu+1} = (r^{\nu}, r^{\nu})/(r^{\nu}, Ar^{\nu}),$$

$$\rho_1 = 1, \quad \rho_{\nu+1} = \left[1 - \frac{\omega_{\nu+1}}{\omega_{\nu}} \frac{(r^{\nu}, r^{\nu})}{(r^{\nu-1}, r^{\nu-1})} \frac{1}{\rho_{\nu}}\right]^{-1}, \quad \nu \geq 1.$$

If r^{ν} is replaced by $b - Ax^{\nu}$ in (3.22), then the three-term form of CGM can be written as follows:

$$x^{\nu+1} = \rho_{\nu+1}\{\omega_{\nu+1}[(I-A)x^{\nu} + b] + (1-\omega_{\nu+1})x^{\nu}\} + (1-\rho_{\nu+1})x^{\nu-1}. \tag{3.23}$$

From (3.23) we know that CGM is a polynomial acceleration method based on the basic iterative method whose iterative matrix is $T = I - A$.

From (3.4) (with $H = I$) we know that, when A in (2.1) is SPSD, A is SPD in the subspace $\mathcal{R}(A)$. Therefore, $\|\cdot\|_{A^{1/2}}$ is a vector norm in $\mathcal{R}(A)$. Let $x^{**} = A^{+}b \in \mathcal{R}(A)$ be a solution of (2.1). It is easy to prove the following theorem.

Theorem 3.1. *Let the initial vector x^0 have a decomposition:*

$$x^0 = x_1^0 + x_2^0, \quad x_1^0 \in \mathcal{N}(A), \quad x_2^0 \in \mathcal{R}(A).$$

Denote

$$x_2^0 + \text{span}\{r^0, Ar^0, \cdots, A^{\nu}r^0\} \equiv x_2^0 + S_{\nu+1} \subset \mathcal{R}(A). \tag{3.24}$$

Then $x^{\nu+1} \in x^0 + S_{\nu+1}$ and $x^{\nu+1} - x_1^0$ minimize $\|x_2 - x^{**}\|_{A^{1/2}}$ $(\forall x_2 \in x_2^0 + S_{\nu+1})$.

In regard to the rate of convergence of CGM we know from (3.10) that the error vector $\mathring{\varepsilon}^{\nu}$ of x^{ν} can be expressed as follows:

$$\mathring{\varepsilon}^{\nu} = x_1^0 + \mathring{Q}_{\nu}(T)\varepsilon_2^0 \equiv x_1^0 + \mathring{\varepsilon}_2^{\nu}.$$

Theorem 3.2. *Assume A is* SPSD. *Then we have the following iteration error bound of the* CGM *for* (2.1)*:*

$$\|\mathring{\varepsilon}_2^{\nu}\|_{A^{1/2}} \leq \frac{2\tilde{r}^{\nu/2}}{1+\tilde{r}^{\nu}}\|\varepsilon_2^0\|_{A^{1/2}}, \tag{3.25}$$

where

$$\tilde{r} = \left(\frac{\sqrt{\mathcal{K}_{\gamma}(A)} - 1}{\sqrt{\mathcal{K}_{\gamma}(A)} + 1}\right)^2, \quad \mathcal{K}_{\gamma}(A) = \frac{M(A)}{m_0(A)}, \tag{3.26}$$

while

$$m_0(A) = \min\{\lambda : \lambda \in \sigma(A)\backslash\{0\}\}. \tag{3.27}$$

Proof. Apply Theorem 3.1 and (3.18).

Corollary. If A is SPSD, then for the error $\mathring{\varepsilon}^{\nu}$ of CGM:

$$\mathring{\varepsilon}^{\nu} = x_1^0 + (x_2^{\nu} - x^{**}),$$

there holds

$$\|x_2^{\nu} - x^{**}\|_{A^{1/2}}^2 \leq 4\left(\frac{\sqrt{\mathcal{K}_{\gamma}(A)} - 1}{\sqrt{\mathcal{K}_{\gamma}(A)} + 1}\right)^{2\nu}\|x_2^0 - x^{**}\|_{A^{1/2}}^2. \tag{3.28}$$

We now discuss the conjugate gradient acceleration based on the basic iterative method (2.2), i.e. the Generalized Conjugate Gradient Method (GCGM).

Since we have assumed that the basic iterative method (2.2) is symmetrizable and (iii) of Theorem 2.1 is satisfied, (2.1) is equivalent to the following preconditioned system:

$$\hat{A}\hat{x} = \hat{b}, \tag{3.29}$$

where

$$\hat{A} = W(I-T)W^{-1}, \quad \hat{x} = Wx, \quad \hat{b} = WHb. \tag{3.30}$$

By applying, CGM (3.21) to (3.29) and making substitutions:

$$\hat{x}^\nu = Wx^\nu, \quad \hat{r}^\nu = W\delta^\nu, \quad \hat{p}^\nu = Wp^\nu, \tag{3.31}$$

where δ^ν is the pseudo-residual vector of x^ν:

$$\delta^\nu = Hb - (I-T)x^\nu = Hr^\nu, \tag{3.32}$$

we have the following GCG Algorithm:

$$\begin{aligned} x^{\nu+1} &= x^\nu + \alpha_\nu p^\nu, \\ p^\nu &= \delta^\nu + \beta_\nu p^{\nu-1} (p^0 = \delta^0 = Hr^0 = H(b - Ax^0)), \\ \alpha_\nu &= \frac{(Wp^\nu, W\delta^\nu)}{(Wp^\nu, W(I-T)p^\nu)} \left[= \frac{(W\delta^\nu, W\delta^\nu)}{(Wp^\nu, W(I-T)p^\nu)} \right], \\ \beta_\nu &= -\frac{(W\delta^\nu, W(I-T)p^{\nu-1})}{(Wp^{\nu-1}, W(I-T)p^{\nu-1})} \left[= \frac{(W\delta^\nu, W\delta^\nu)}{(W\delta^{\nu-1}, W\delta^{\nu-1})} \right], \end{aligned} \tag{3.33}$$

or the equivalent three-term recurrence scheme (cf. (3.22)):

$$x^{\nu+1} = \rho_{\nu+1}(\omega_{\nu+1}\delta^\nu + x^\nu) + (1-\rho_{\nu+1})x^{\nu-1}, \quad \delta^\nu = Hr^\nu, \tag{3.34}$$

where

$$\begin{aligned} \omega_{\nu+1} &= \left[1 - \frac{(W\delta^\nu, WT\delta^\nu)}{(W\delta^\nu, W\delta^\nu)}\right]^{-1}, \\ \rho_1 = 1, \quad \rho_{\nu+1} &= \left[1 - \frac{\omega_{\nu+1}}{\omega_\nu} \frac{(W\delta^\nu, W\delta^\nu)}{(W\delta^{\nu-1}, W\delta^{\nu-1})} \frac{1}{\rho_\nu}\right]^{-1}, \quad \nu \geq 1. \end{aligned} \tag{3.35}$$

By applying Theorem 3.2 we can prove the following theorem.

Theorem 3.3. *For the error vector ε^ν of x^ν produced by* GCGM:

$$\varepsilon^\nu = x_1^0 + (x_2^\nu - x^{**}) \equiv x_1^0 + \varepsilon_2^\nu$$

there holds

$$\|\varepsilon_2^\nu\|_{[W^TW(I-T)]^{1/2}} \leq \frac{2\tilde{r}^{\nu/2}}{1+\tilde{r}^\nu} \|\varepsilon_2^0\|_{[W^TW(I-T)]^{1/2}}, \tag{3.36}$$

where

$$\tilde{r} = \frac{1-\sqrt{1-\tilde{\sigma}^2}}{1+\sqrt{1-\tilde{\sigma}^2}}, \quad \tilde{\sigma} = \frac{M_\gamma(T) - m_\gamma(T)}{2 - M_\gamma(T) - m_\gamma(T)}. \tag{3.37}$$

§4. Conjugate Gradient Acceleration for the CGW Method

Let A in (2.1) be (not necessarily symmetric) positive semidefinite. Make the CGW splitting[2,7,5] :

$$A = M - N \equiv \frac{A + A^T}{2} - \frac{A^T - A}{2}, \tag{4.1}$$

where M is SPSD and N is asymmetric. The proof of the following lemma is obvious and so is omitted.

Lemma 4.1. *Assume that A in (2.1) is positive semidefinite* (PSD) *and*

$$\mathcal{N}(A + A^T) = \mathcal{N}(A) \cap \mathcal{N}(A^T). \tag{4.2}$$

Then

$$\mathcal{R}(A), \quad \mathcal{R}(A^T) \subset \mathcal{R}(M). \tag{4.3}$$

In this section we always assume that A is PSD and satisfies (4.2).

With reference to (3.34) for the symmetrizable basic iterative method we construct the following iterative scheme:

$$\begin{aligned} x^{\nu+1} &= \rho_{\nu+1}(\omega_{\nu+1}\delta^\nu + x^\nu) + (1 - \rho_{\nu+1})x^{\nu-1}, \quad \rho_1 = 1, \\ \delta^\nu &= M^+ r^\nu = M^+(b - Ax^\nu). \end{aligned} \tag{4.4}$$

Theorem 4.1. *If the parameters $\{\rho_\nu, \omega_\nu\}$ are chosen such that*

$$\omega_\nu = 1, \quad \rho_1 = 1, \quad \rho_{\nu+1} = \left[1 + \frac{(\delta^\nu, M\delta^\nu)}{(\delta^{\nu-1}, M\delta^{\nu-1})}\frac{1}{\rho_\nu}\right]^{-1}, \quad \nu \geq 1, \tag{4.5}$$

then the sequence $\{\delta^\nu\}$ produced by (4.4) satisfies the orthogonal relations:

$$(\delta^i, M\delta^j) = 0, \quad i \neq j.$$

Proof. From (4.4) we get

$$M\delta^{\nu+1} = M\delta^{\nu-1} - \rho_{\nu+1}[\omega_{\nu+1}A\delta^\nu + M(\delta^{\nu-1} - \delta^\nu)]. \tag{4.6}$$

From Lemma 4.1 we know

$$\delta^\nu = M^+ r^\nu \in \mathcal{R}(M).$$

Since M is SPD in subspace $\mathcal{R}(M)$, the conclusion of the theorem can be deduced by induction.

(4.4) and (4.5) form the GCG Algorithm for the CGW method.

We now consider the convergence and the iteration error bound of the GCGM. Let

$$A = M(I - T), \text{ i.e. } \quad T = M^+ N, \quad \delta^0 = M^+ r^0 \tag{4.7}$$

and

$$S_\nu = \text{ span } \{\delta^0, T\delta^0, \cdots, T^{\nu-1}\delta^0\} \subset \mathcal{R}(M). \tag{4.8}$$

From (4.6) and the M-orthogonality of $\{\delta^i\}$ we have

$$S_\nu = \text{ span } \{\delta^0, \delta^1, \cdots, \delta^{\nu-1}\}. \tag{4.9}$$

From (4.4) and (4.5) we have

$$x^{\nu} \in x^0 + S_{\nu} = x_1^0 + x_2^0 + S_{\nu}, \tag{4.10}$$

where

$$x_1^0 \in \mathcal{N}(A), \quad x_2^0 \in \mathcal{R}(A^T) \subset \mathcal{R}(M). \tag{4.11}$$

Then x^{ν} can be expressed as follows:

$$x^{\nu} = x_1^0 + x_2^{\nu}, \quad x_1^0 \in \mathcal{N}(A), \quad x_2^{\nu} \in \mathcal{R}(M).$$

Let $x^{**} = A^+ b \in \mathcal{R}(A^T) \subset \mathcal{R}(M)$. Then we have

$$\varepsilon^{\nu} = x^{\nu} - x^{**} = x_1^0 + (x_2^{\nu} - x^{**}) \equiv x_1^0 + \varepsilon_2^{\nu}. \tag{4.12}$$

Since

$$M\delta^{\nu} = r^{\nu} = -A(x^{\nu} - x^{**}) = -A(x_2^{\nu} - x^{**}) = -A\varepsilon_2^{\nu} \tag{4.13}$$

and since the sequence $\{\delta^i\}$ is M-orthogonal, we have

$$(\delta, A\varepsilon_2^{\nu}) = 0, \quad \forall \delta \in S_{\nu}. \tag{4.14}$$

From (4.4) and (4.13) we get

$$\varepsilon_2^{\nu+1} = (1 - \rho_{\nu+1})\varepsilon_2^{\nu-1} + \rho_{\nu+1}(\delta^{\nu} + \varepsilon_2^{\nu}) = (1 - \rho_{\nu+1})\varepsilon_2^{\nu-1}$$
$$-\rho_{\nu+1}(M^+ A\varepsilon_2^{\nu} - \varepsilon_2^{\nu}) = \rho_{\nu+1} T\varepsilon_2^{\nu} + (1 - \rho_{\nu+1})\varepsilon_2^{\nu-1}.$$

Hence

$$\varepsilon_2^{\nu} = p_{\nu}(T)\varepsilon_2^0, \quad p_{\nu}(\mu) \in P_{\nu}, \tag{4.15}$$

and $p_{\nu}(\mu)$ is an odd (even) polynomial when ν is odd (even). From (4.15) we have

$$x_2^{\nu} = x^{**} + p_{\nu}(T)\varepsilon_2^0. \tag{4.16}$$

Now it is easy to derive some variational properties of $\{x_2^{\nu}\}$. Notice that M is SPD in $\mathcal{R}(M)$. The proofs of the following lemmas and theorems can be made with reference to those in [4]. In what follows $\{x^{\nu}\}$ is the sequence produced by GCGM (4.4) and (4.5).

Lemma 4.2. *Let $x^{2\nu} = x_1^0 + x_2^{2\nu}$. Then $x_2^{2\nu} \in x_2^0 + (I+T)S_{2\nu}$ and*

$$\|\varepsilon_2^{2\nu}\|_{M^{1/2}} \equiv \|x_2^{2\nu} - x^{**}\|_{M^{1/2}} = \min\{\|y - x^{**}\|_{M^{1/2}} : \quad y \in x_2^0 + (I+T)S_{2\nu}\}.$$

Lemma 4.3. *Let $x^{2\nu+1} = x_1^0 + x_2^{2\nu+1}$. Then $x_2^{2\nu+1} \in x_2^0 + \delta^0 + (I+T)S_{2\nu+1}$ and*

$$\|\varepsilon_2^{2\nu+1}\|_{M^{1/2}} \equiv \|x_2^{2\nu+1} - x^{**}\|_{M^{1/2}} = \min\{\|y - x^{**}\|_{M^{1/2}} : \quad y \in x_2^0 + \delta^0 + (I+T)S_{2\nu+1}\}.$$

From the optimal approximation properties of $\{x_2^{\nu}\}$ described by Lemma 4.2 and Lemma 4.3 we get the following theorem.

Theorem 4.2. *Let $q_{\nu}(\mu)$ be any polynomial with degree $\le \nu$ and satisfy the conditions: $q(1) = 1$ and $q(-1) = (-1)^{\nu}$. Then*

$$\|\varepsilon_2^{\nu}\|_{M^{1/2}} = \|x_2^{\nu} - x^{**}\|_{M^{1/2}} \le \|q_{\nu}(T)\ (x_2^0 - x^{**})\|_{M^{1/2}}. \tag{4.17}$$

By Theorem 4.2 and the choice of a proper sequence of polynomials $\{q_\nu(\mu)\}$ we can get an error bound of the GCGM based on the CGW splitting.

Since M is SPD in $\mathcal{R}(M), T = M^+N$ is asymmetric in the inner product $(\cdot,\cdot)_M$. Let T_M denote T restricted to $\mathcal{R}(M)$ and write

$$\Lambda = \rho(T_M) = \|T_M\|_{M^{1/2}}. \tag{4.18}$$

Then we can get the following theorem:

Theorem 4.3. *Let A in* (2.1) *be positive semidefinite and satisfy* (4.2). *Then we have the error bound for the* GCGM (4.4)–(4.5) *based on the* CGW *splitting* (4.1) *of A:*

$$\|\varepsilon_2^\nu\|_{M^{1/2}} \le \frac{2\tilde{r}^{\nu/2}}{1+(-1)^\nu \tilde{r}^\nu}\|\varepsilon_2^0\|_{M^{1/2}}, \tag{4.19}$$

where

$$\tilde{r} = \frac{\sqrt{1+\Lambda^2}-1}{\sqrt{1+\Lambda^2}+1}. \tag{4.20}$$

References

[1] A. Berman and R.J. Plemmons, Non-negative Matrices in the Mathematical Sciences, Academic Press, New York, 1979.

[2] P. Concus and G.H. Golub, A generalized conjugate gradient method for nonsymmetric systems of linear equations, in Lecture Notes in Economics and Mathematical Systems 134, 56–65 (R. Glowinski and J.L. Lions, eds.), Springer-Verlag, New York, 1976.

[3] P. Concus, G.H. Golub and D.P. O'Leary, A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, in Sparse Matrix Computation (J.R. Bunch and D.J. Rose, eds.), 309–332, Academic Press, New York, 1976.

[4] S.C. Eisenstat, A note on the generalized conjugate gradient method, *SIAM J. Numer. Anal.*, **20** (1983), 358–361.

[5] L.A. Hageman and D.M. Young, Applied Iterative Methods, Academic Press, New York, 1981.

[6] A. Ralston and H.S. Wilf, Mathematical Methods for Digital Computers, John Wiley & Sons, New York, 1960.

[7] O. Widlund, A Lanczos method for a class of nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.*, **15**(1978), 801–812.

[8] G.I. Marchuk and Yu.A. Kuznetsov, Iterative methods and quadratic functionals, Nauka, Novosibirsk, 1975. (in Russian)

[9] Z.H. Cao, The Matrix Eigenvalue Problem, Shanghai Science & Technology Press, 1980. (in Chinese)